

Cost effective, robust estimation of measurement uncertainty from sampling using unbalanced ANOVA

Article (Accepted Version)

Rostron, Peter and Ramsey, Michael (2012) Cost effective, robust estimation of measurement uncertainty from sampling using unbalanced ANOVA. Accreditation and Quality Assurance, 17 (1). pp. 7-14. ISSN 1432-0517

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/41740/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Cost effective, robust estimation of measurement uncertainty from sampling using unbalanced ANOVA

Peter D. Rostron

School of Life Sciences, University of Sussex, Falmer, Brighton, BN1 9QG, UK

Tel: 01273 872958

email: pr52@sussex.ac.uk

Michael H. Ramsey

School of Life Sciences, University of Sussex, Falmer, Brighton, BN1 9QG, UK

Abstract

There is an increasing appreciation that the uncertainty in environmental measurements is vitally important for their reliable interpretation. However, the adoption of methods to estimate this uncertainty has been limited by the extra cost of implementation. One method that can be used to estimate the random components of uncertainty in the sampling and analytical processes requires the collection of duplicate samples at 10% of the primary sampling locations, and duplicating the analyses of these samples. A new program has been written and applied to a modified experimental design to enable a 33% reduction in the cost of analysing this 10% subset, whilst accommodating outlying values. This unbalanced robust analysis of variance (U-RANOVA) uses an *unbalanced* rather than the *balanced* experimental design usually employed. Simulation techniques have been used to validate the results of the program, by comparison of the results between the proposed unbalanced and the established balanced designs. Comparisons are also made against the seed parameters (mean and standard deviation) used to simulate the parent population, prior to the addition of a proportion (up to 10%) of outlying values. Application to a large number of different simulated populations shows that U-RANOVA produces results that are effectively indistinguishable from the results produced by the accepted balanced approach, and are equally close to the true (seed) parameters of the parent normal population.

Keywords

Uncertainty, unbalanced design, duplicate sample, robust ANOVA, accommodating outliers, optimized uncertainty.

Introduction

Characterisation of the intensity and distributions of analytes within a bulk of material usually requires the taking of a number of samples from discrete points within the parent volume, where each sample represents a sampling target within the parent volume. This is because it is rarely possible to analyse the whole parent volume, for example it would be impractical to analyse the entire soil mass in an area of contaminated land. These primary samples are either analysed *in situ* or removed, processed, and analysed *ex situ*, in order to produce measurements which are intended to be representative of their respective sampling targets within the parent volume [1].

These measurements are subject to uncertainties, which arise both at the sampling and at the analytical stages of the investigation. Uncertainties result from both systematic and random errors caused by the methods used. As the magnitudes of these errors affect the quality of the measurements, it is important to estimate these uncertainties before basing decisions on the characterisation of a parent volume (e.g. a plan for land remediation). Ramsey *et al.* [2] provides an overview of methods that can be used for the estimation of systematic and random errors in both the sampling and the analytical processes. Of the four methods for the estimation of the random component of uncertainty described in that review, the simplest is identified as the duplicate method, which uses a *balanced*

experimental design. In this design, a number of duplicate samples are taken (from 10% of the primary sampling locations, and a minimum of 8), and each of them chemically analysed twice (Fig. 1).

[Fig 1]

Fig. 1 The balanced experimental design, which can be used to estimate the sampling and analytical components of measurement uncertainty using the duplicate method. Two samples are taken at each primary sampling point, and each sample is chemically analysed twice [1]

When the analysis of primary samples is performed in order to characterise a number of sampling targets within a parent volume, the resultant set of measurements contains three components of variance: variance due to the actual variation of the particular property being measured (e.g. analyte concentration), between the sampling targets (the between-target variance); variance due to uncertainty in the sampling method; and variance due to uncertainty in the analytical method. The first of these, the between-target variance, is the particular component of interest for characterization, as this is a parameter of the distribution of true values of the analyte within the sampling target. Therefore there is a need to separate this component from the total variance in the measurement set [2]. Garrett [3] suggests that for geochemical data, where economic interests often rely on subtle changes in geochemistry, the total variance in the data should exceed the combined sampling and analytical variance by a factor of at least 4. When this is not the case, there is a greater than 5% chance that observed variability could be due to variances inherent in the sampling and analytical processes.

These three components can be separated by the use of classical analysis of variance (ANOVA), which has been in use for determining the significance of areal variation in geochemical datasets since the 1960s. A drawback of the balanced design in geochemical surveys is that of increased cost, especially when several levels of variability are required. One way in which this cost can be reduced is to use the *simplified* design (not illustrated) as quoted in the Eurachem guide [1], where only one analysis is performed on each duplicated sample, yielding an estimate of total uncertainty. If required, the sampling uncertainty can then be estimated by subtracting an external estimate of the analytical uncertainty from this total. An alternative approach is to modify the system of duplication so that each component of the variability is duplicated once only [4]. Termed the *unbalanced* design, this design is also identified as being of potential use in the Eurachem guide [1]. The three-tier experimental design already described (Fig. 1) can be modified to an unbalanced design by removing one analysis operation from one of the samples from each duplicated primary sample (Fig. 2).

As the unbalanced design requires only 2 additional analyses per duplicated sample, instead of an additional 3 as for the balanced design, using the unbalanced design reduces the cost of analysing the 10% subset of replicate locations by 33%. As an example, a survey in which 100 primary samples are taken and chemically analysed, eight of which are designated as duplicate primary samples, would require the analytical procedure to be carried out a total of 124 times (i.e. 100 single samples + 8x3 duplicates) using the balanced design. This number would reduce to 116 (100 single samples + 8x2 duplicates) for the unbalanced design, equating to a saving of 6.5% on the overall cost of analysis. This could be a significant saving particularly if the costs of chemical analysis were high.

[Fig 2]

Fig. 2 The unbalanced experimental design, where only one of the two samples from a duplicated primary sample undergoes duplicate analysis. This reduces the total number of chemical analyses required [1]

It has been found in trials that 8 duplicate primary samples is typically a minimum number for the estimation of uncertainty by the duplicate method. Although increasing the number of these duplicates beyond 8 reduces the confidence interval on the uncertainty estimates, the marginal improvement so obtained may not justify the increased costs [5]. Performing the balanced design on 8 duplicate primary samples results in 8 sample duplicates and 16 analytical duplicates, whereas using the unbalanced design produces 8 sample duplicates and 8 analytical duplicates. One characteristic of the balanced design therefore is that there are fewer degrees of freedom in the estimation of sampling uncertainty than in the estimation of analytical uncertainty. Hence the confidence interval on the estimate of sampling uncertainty is larger than that on the estimate of analytical uncertainty. In the *unbalanced* design, the numbers of degrees of freedom on the sampling and analytical components of uncertainty are made more equal. As these components may be considered to be of equal importance, this enables an equal amount of effort to be applied to each level, instead of twice the effort being made to estimate analytical uncertainty as is made to estimate sampling uncertainty, which is the case for the balanced design. Therefore this method potentially enables a more efficient allocation of resources.

One of the assumptions of classical ANOVA is that the distribution of errors within each level of variance approximates to a Gaussian distribution. However, data from surveys (e.g. geochemical) often contain a small number of outlying values (i.e. values that are untypically far from the mean) and distributions may be heavily tailed [6,7]. When classical statistics are used such outliers result in a biased mean and high standard deviation, which are not good representations of the main body of the data. A traditional approach to dealing with such outliers is to employ statistical significance tests in order to decide whether particular outliers should be excluded from the dataset [7]. This method, however, leads to an underestimation of the variance achievable by a particular analytical method [8], and the same applies to sampling variance. An alternative approach is to use robust statistics [7]. These methods accommodate, rather than reject, outliers, resulting in estimators of central tendency (e.g. the mean or median) and estimators of the variability in the data (e.g. variance or standard deviation) that are relatively unaffected by small populations of outliers [6]. A number of different approaches to robust estimations of these parameters exist, e.g. those given by Rousseeuw & Verboven for very small datasets [9]. The methodology used in this current work is an iterative approach that can only practically be performed by a computer program. The robust mean μ_r is initially estimated as the classical mean, and the robust standard deviation σ_r as the median of the absolute differences between duplicated measurements. Any values that are found to exceed $\mu_r + c \sigma_r$ are replaced with $\mu_r + c \sigma_r$, and any values that fall below $\mu_r - c \sigma_r$ are replaced with $\mu_r - c \sigma_r$, where c is a factor between 1 and 2 (typically, as here, set to 1.5). The robust statistics μ_r and σ_r are then recalculated, and the process repeated multiple times, until μ_r stabilizes (converges) at an acceptable level of accuracy [7,8].

The original robust ANOVA program for geochemical surveys (ROBCOOP) was based on a program listing published by the Analytical Methods Committee and uses the balanced design (Fig. 1) [8]. It was tested using simulated datasets, and the estimated robust means and standard deviations produced were shown to be very close to the seed distribution parameters used to create the simulated data. In contrast, classical estimates of these statistics were found to differ from the seed parameters by up to 1 order of magnitude in some cases [6]. This FORTRAN program has since been partly re-written in Microsoft Visual Basic for Applications (Excel) so that it can be incorporated into Excel utilities. The new program RANOVA has been produced for the specific case of a two-stage nested design using the duplicate method, as quoted in the Eurachem guide [1]. It has been shown to produce identical results to the program ROBAN, which is also based on ROBCOOP, and is available free of charge from the Royal Society of Chemistry website.

However, no method has previously been devised for performing *robust* ANOVA on duplicate measurements obtained using the unbalanced design. Consequently the program RANOVA has been modified to perform robust ANOVA on the unbalanced design (Fig. 2). As this reduces the number of additional measurements per duplicate primary sampling location from 3 to 2, this enables a 33% reduction in the total costs of analysing the 10% subset of replicate locations in order to estimate measurement uncertainty, as discussed above. The modified program has been named U-RANOVA. The aim of the following experiments was to verify that this program produces estimates

of the robust mean and component standard deviations that are approximately equivalent to the robust statistics produced by ROBCOOP and RANOVA, and also to the seed parameters used to construct simulated datasets.

The objectives of this paper are as follows:

1. *Explain the advantages of using an unbalanced design for the empirical estimation of the random component of measurement uncertainty that arises from sampling.*
2. *Describe and validate a new computer program that can estimate uncertainty for population data with up to 10% of outlying values using the unbalanced design.*

Methods

The new program U-RANOVA was tested in two stages. *Test 1* compared the variances estimated by U-RANOVA with the variances obtained by ROBCOOP the 1992 study, using the same data as that study [6]. *Test 2* was performed on newly generated populations, comprising simulations of analyte concentration values. In this case the estimated variances were compared with the robust estimates made by the program RANOVA for balanced survey designs, as well as with the seed parameter values.

Test 1

The original FORTRAN program for robust ANOVA of balanced experimental designs (ROBCOOP) was tested using 4 simulated populations, to which outlying values were subsequently added as explained below. Each of the 4 populations was initially produced using the same seed parameters (Table 1).

These populations, designated A, B, C & D, each represent measurements of 100 sampling targets from a parent volume, with duplicated samples and duplicated analysis for each sampling target. Thus they comprise four columns of numbers, simulating sets of measurements that might be taken using a balanced experimental design (Fig. 1). In order to test the effects that outliers had on the estimations of variance, population B had 10% of the analytical duplicates (10% of column 4) overwritten by simulated high values. Thus 2.5% of the total population were set high, meaning that 5% of the differences between analytical duplicates were replaced by outlying values in the ANOVA calculation. In the same way, population C had 10% of the sampling duplicates (columns 3 & 4) overwritten with high values (5% of the population), and population D had 10% of the between-target values (all four columns, 10% of the population) set high [6]. An example showing simulated measurements for 15 sampling targets (including outliers) is shown in Fig 3. The original data were available to the author, and so could be input into U-RANOVA, and the estimated variances produced compared with the published results. As all four populations contained four columns of simulated measurements, and the unbalanced experimental design generates just three measurements per site (Fig. 2), three columns were chosen for the test on each population. The columns were chosen to ensure that the simulated outlying values were included in the U-RANOVA estimations [6].

Table 1 Seed distribution parameters of the 4 simulated populations used in the 1992 study [6]

Mean	100 $\mu\text{g g}^{-1}$
s (Analytical)	2 $\mu\text{g g}^{-1}$
s (Sampling)	5 $\mu\text{g g}^{-1}$
s (Between-target)	10 $\mu\text{g g}^{-1}$

s = standard deviation

Test 2

A more comprehensive set of tests was performed with the intention of comparing component standard deviations estimated by U-RANOVA with those produced by RANOVA. In total, 33 trials were performed, based on populations produced from a combination of three sets of seed distribution parameters (Table 2) and eleven outlier scenarios (Table 3). The three seed parameter combinations (Table 2) were chosen based on the following reasoning: Seed 1 is a repeat of the seed parameters used in the 1992 study (Table 1); Seed 2 equalises the analytical and sampling standard deviation at $5 \mu\text{g g}^{-1}$; Seed 3 uses parameters that are intended to be reasonable representations of the magnitude of variances that might be found during an investigation with a very high proportion of sampling variance (e.g. a contaminated land area) [1].

Table 2 Seed distribution parameters used in *Test 2*

	Seed 1	Seed 2	Seed 3
Mean	$100 \mu\text{g g}^{-1}$	$100 \mu\text{g g}^{-1}$	$100 \mu\text{g g}^{-1}$
s (Analytical)	$2 \mu\text{g g}^{-1}$	$5 \mu\text{g g}^{-1}$	$5 \mu\text{g g}^{-1}$
s (Sampling)	$5 \mu\text{g g}^{-1}$	$5 \mu\text{g g}^{-1}$	$30 \mu\text{g g}^{-1}$
s (Between-target)	$10 \mu\text{g g}^{-1}$	$10 \mu\text{g g}^{-1}$	$50 \mu\text{g g}^{-1}$

s = standard deviation

For each of the seed parameter combinations in Table 2, MS-Excel was used to randomly generate 1000 normally distributed “base” population of simulated measurements, each intended to simulate 4 measurements (see Fig 1) of each of 100 sampling targets. An additional ten simulated populations were then generated from each base population according to the outlier types and adjustments described in Table 3. In order to better represent measurements that might be obtained during real-life experiments, outliers were randomly distributed through each population. An example extract from a population with combined analytical, sampling and between-target outliers is shown in Fig 4. The 33 000 randomly generated test populations eventually generated were retained on computer disk and are available from the authors on request.

[Fig 3]

Fig. 3 Data extracts from the four trial populations used by Ramsey *et al.* [6]. For each trial the numbers shown represent measurements in $\mu\text{g g}^{-1}$ of 15 out of a total of 100 sampling targets. The numbers were generated in such a way as to simulate measurements that might have been obtained using the balanced design (Fig 1). The boxed numbers are the outlying values, generated by adding a fixed number to the base population shown in Trial A.

Table 3 The 4 different outlier types and 3 different outlier adjustments that were applied to each base population in *Test 2*, generating a total of 10 additional populations with outliers for each base population.

Outlier	Outlier	Description of outlier adjustment procedure
---------	---------	---

Type	adjustments ($\mu\text{g g}^{-1}$)	
5% analytical	-90	10 different sampling targets selected at random throughout population. For each selected target, analytical outlier randomly assigned to either Sample 1 or Sample 2 with equal probability (Fig 2). Outlier adjustment then applied to either Analysis 1 or Analysis 2 of selected sample with equal probability.
	+100	
	+200	
10% sampling	-90	10 different sampling targets selected at random throughout population. For each selected target, outlier randomly assigned to either Sample 1 or Sample 2 (Fig1). Outlier adjustment applied to Analysis 1 and Analysis 2 of selected sample.
	+100	
	+200	
10% between- target	-90	10 different sampling targets selected at random throughout population. For each selected target, outlier adjustment applied to Analysis 1 and Analysis 2 of both samples (Fig 1).
	+100	
	+200	
10% combined	+100	3 between-target outlier adjustments applied as described above (Outlier Type = 10% between-target), followed by 3 sampling outlier adjustments applied as above (Outlier Type = 10% sampling), followed by 4 analytical outlier adjustments as described above (Outlier Type = 10% analytical).

RANOVA and U-RANOVA were used to estimate the robust statistics for each population. The mean and component standard deviations were then averaged across the 1000 populations for each of the 33 trials, thus producing an estimate of the bias between the results of the balanced and the unbalanced designs in each case.

Results

Test 1

Comparisons of the variances estimated by U-RANOVA, expressed as standard deviations, are found to be within 6% of the results originally obtained using robust ANOVA on a balanced design (Table 4). Identical values would not be expected as the estimates for the unbalanced design were made on three columns of data, whereas the estimates for the balanced design were made on all four columns. The U-RANOVA estimates are also better representations of the seed population parameters (Table 1) than were obtained by classical ANOVA techniques, the most extreme difference being the estimate of $6.02 \mu\text{g g}^{-1}$ for the sampling standard deviation of population D. This is 20% higher than the seed parameter of $5 \mu\text{g g}^{-1}$. In comparison, the standard deviations estimated by Ramsey *et al.* (1992) using classical ANOVA differed by one order of magnitude in some cases [6].

[Fig 4]

Fig. 4 Data extracts from a simulated population with analytical, sampling and between-target outliers. The boxed numbers ($\mu\text{g g}^{-1}$) are the outlying values, generated by adding a fixed number (100) to the base population.

Table 4 Comparison of results of ROBUST ANOVA on an unbalanced experimental design (U-RANOVA), with previously published results for a balanced design (Ramsey *et al.*), showing that the differences between the estimated component standard deviations are <6%. The differences between estimates are shown as a percentage of the original estimate from the 1992 study [6]

Original population ID	Data source	Mean ($\mu\text{g g}^{-1}$)	Standard Deviations ($\mu\text{g g}^{-1}$)		
			Analytical (2.5% population outliers)	Sampling (5% population outliers)	Between Target (10% population outliers)
A (no outliers)	Ramsey <i>et al.</i> (1992)	100.13	1.92	5.24	9.62
	U-RANOVA (2011)	100.17	2.02	5.25	9.51
	Difference %	0.04	5.21	0.19	-1.14
B (analytical outliers)	Ramsey <i>et al.</i> (1992)	102.00	2.05	5.80	10.87
	U-RANOVA (2011)	102.30	2.02	5.85	11.17
	Difference %	0.29	-1.46	0.86	2.76
C (sampling outliers)	Ramsey <i>et al.</i> (1992)	102.67	1.92	5.82	11.88
	U-RANOVA (2011)	102.30	2.02	5.85	11.17
	Difference %	-0.36	5.21	0.52	-5.98
D (between-target outliers)	Ramsey <i>et al.</i> (1992)	102.28	2.23	6.02	11.83
	U-RANOVA (2011)	102.31	2.34	5.97	11.92
	Difference %	0.03	4.93	-0.83	0.76

Test 2

Comparison of the output from RANOVA and U-RANOVA on the 33 seed parameter/outlier type combinations (Table 5), show that the average differences between the standard deviations for balanced and unbalanced designs are comparatively small, with all showing differences of <7% averaged over 1000 trials. The maximum differences of -6.1% are found in the sampling standard deviations for Seed 2. Some differences would be expected as the balanced design is using 4 columns of data whereas the unbalanced design is using only 3, and so the unbalanced design omits some of the randomly placed analytical outliers. The robust means are all very good approximations, with the means for the unbalanced design being within 1.0% of the means for the balanced design. The means and component standard deviations are again much better estimates of the seed population parameters than are the classical ANOVA results obtained in the 1992 study [6]. In this case, the maximum differences are in the sampling standard deviations of the populations where sampling outliers have been added to Seed 2, which at $6.6 \mu\text{g g}^{-1}$ are 32% higher than the seed population parameter of $5 \mu\text{g g}^{-1}$ (Table 2). That Seed 2 yields the highest percentage differences, both when the unbalanced design is compared to the balanced design, and also when the standard deviations are compared to the seed parameters, indicate that this method may not be optimal when the seed analytical and sampling standard deviations are equal in magnitude. Further experiments, where the seed analytical standard deviations have been set higher than the sampling standard deviations, have also shown progressively larger differences between estimates of standard deviations obtained from the unbalanced and the balanced designs. For example, in one case where the seed analytical standard deviation exceeded the sampling standard deviation by a factor of 4, the estimate of sampling standard deviation made by U-RANOVA was 18% higher than that made by RANOVA, when averaged over 1000 trials. In most cases, e.g. chemical contamination of land areas, we consider it unlikely that the variance due to analytical uncertainty will often exceed that due to sampling uncertainty.

Conclusion

The unbalanced experimental design as described in the Eurachem guide [1], can be used in the implementation of the duplicate method to estimate the random component of measurement uncertainty. However, it has not previously been possible to obtain these estimates using robust statistical methods, which accommodate a small proportion (<10% of sampling targets) of outlying values. A computer program U-RANOVA has been written to estimate the robust mean and component variances from data produced by an unbalanced design. This enables the random components of uncertainty to be estimated with fewer analytical measurements than are required by the balanced design. U-RANOVA was validated by inputting population data used in a previously published study, and comparing its output with that obtained in that study (Test 1) [6]. Additional trials were undertaken on datasets containing 1000 unique populations per set, and with 10 different outlier types (Test 2). Results of *Test 1* and averaged results from *Test 2* showed that estimates of population parameters from the balanced and unbalanced designs were much more representative of the seed population parameters than were produced by classical ANOVA [6]. In both cases, the majority of estimates of component standard deviations from the unbalanced experimental design were found to be within 5% of the estimates from the corresponding balanced design, and all were within 7%. This demonstrates that the unbalanced experimental design can be used to obtain robust estimates of uncertainty with a 33% reduction in the cost of analysing the 10% subset of replicate locations required by the duplicate method.

The program U-RANOVA, written in Visual Basic for Applications (Excel) is available at the Royal Society of Chemistry website.

[Table 5]

Table 5: Comparison of the means and standard deviations (SD) estimated by robust ANOVA for unbalanced (U-RANOVA) and balanced (RANOVA) experimental designs. The difference between estimates is shown as a percentage of the estimate calculated by the balanced design. Each difference is the average difference for 1000 simulated populations. Seed parameters are shown in Table 2.

Total standard deviation is the square root of the sum of the component variances,
e.g. $\sqrt{s^2(\text{Analytical}) + s^2(\text{Sampling}) + s^2(\text{Between-target})}$

Acknowledgments

The authors would like to thank Prof. Tom Fearn of University College London, for his helpful advice on certain aspects of the robust calculations.

References

1. Eurachem/CITAC (2007) *Measurement uncertainty arising from sampling*. In: M.H. Ramsey SLRE (ed) Eurachem / CITAC guide
2. Ramsey MH (1998) *Sampling as a source of measurement uncertainty: Techniques for quantification and comparison with analytical sources*. *Journal of Analytical Atomic Spectrometry* 13:97-104
3. Garrett RG (1969) *The determination of sampling and analytical errors in exploration geochemistry*. *Economic Geology* 64:568-569
4. Garrett RG, Goss TI (1980) *Uanova: A fortran iv program for unbalanced nested analysis of variance*. *Computers & Geosciences* 6 (1):35-60
5. Lyn JA, Ramsey MH, Coad DS, Damant AP, Wood R, Boon KA (2007) *The duplicate method of uncertainty estimation: Are eight targets enough?* *Analyst* 132:1147-1152

6. Ramsey MH, Thompson M, Hale M (1992) Objective evaluation of precision requirements for geochemical analysis using robust analysis of variance. *Journal of geochemical exploration* 44:23-36
7. AMC (2001) Robust statistics: A method of coping with outliers. *Analytical Methods Committee AMC tech brief* 6:1-2
8. AMC (1989) Robust statistics - how not to reject outliers. Part1, basic concepts. *Analyst* 114:1693-1697
9. Rousseeuw PJV, Sabine (2002) Robust estimation in very small samples. *Computational Statistics and Data Analysis* (40):741-758